(http://www.pieriandata.com)

Pandas Data Visualization Exercise

This is just a quick exercise for you to review the various plots we showed earlier. Use **df3** to replicate the following plots.

In [6]:

```
import pandas as pd
import matplotlib.pyplot as plt
df3 = pd.read_csv('df3.csv')
%matplotlib inline
```

In [7]:

```
df3
```

Out[7]:

	a	b	c	d
0	0.336272	0.325011	0.001020	0.401402
1	0.980265	0.831835	0.772288	0.076485
2	0.480387	0.686839	0.000575	0.746758
3	0.502106	0.305142	0.768608	0.654685
4	0.856602	0.171448	0.157971	0.321231
...	...	...	...	...
495	0.528705	0.226122	0.055835	0.131962
496	0.324730	0.215201	0.935302	0.794115
497	0.118036	0.264574	0.629206	0.824062
498	0.227021	0.660209	0.851353	0.478676
499	0.466157	0.753000	0.115391	0.279712

500 rows × 4 columns

In [8]:

```
df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 4 columns):
a    500 non-null float64
b    500 non-null float64
c    500 non-null float64
d    500 non-null float64
dtypes: float64(4)
memory usage: 15.8 KB
```

In [9]:
```
df3.head()
```

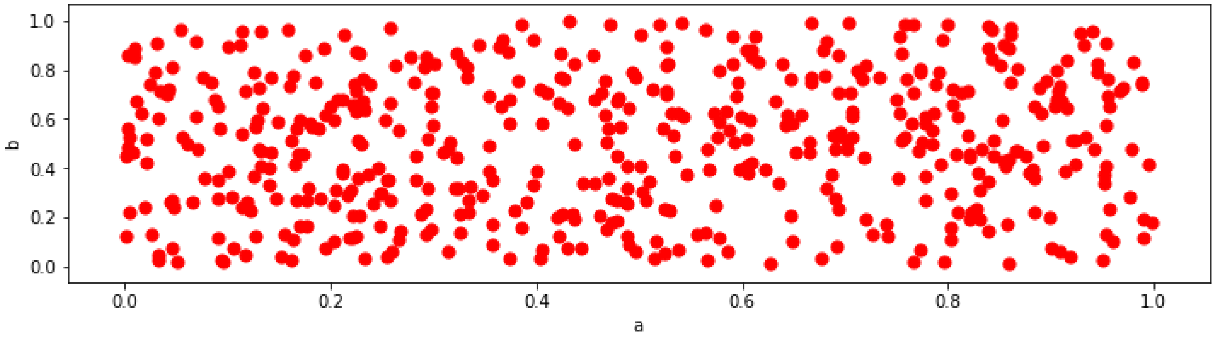
Out[9]:

	a	b	c	d
0	0.336272	0.325011	0.001020	0.401402
1	0.980265	0.831835	0.772288	0.076485
2	0.480387	0.686839	0.000575	0.746758
3	0.502106	0.305142	0.768608	0.654685
4	0.856602	0.171448	0.157971	0.321231

** Recreate this scatter plot of b vs a. Note the color and size of the points. Also note the figure size. See if you can figure out how to stretch it in a similar fashion. Remeber back to your matplotlib lecture...**

In [10]:
```
df3.plot.scatter(x='a',y='b',c='red',s=50,figsize=(12,3))
```

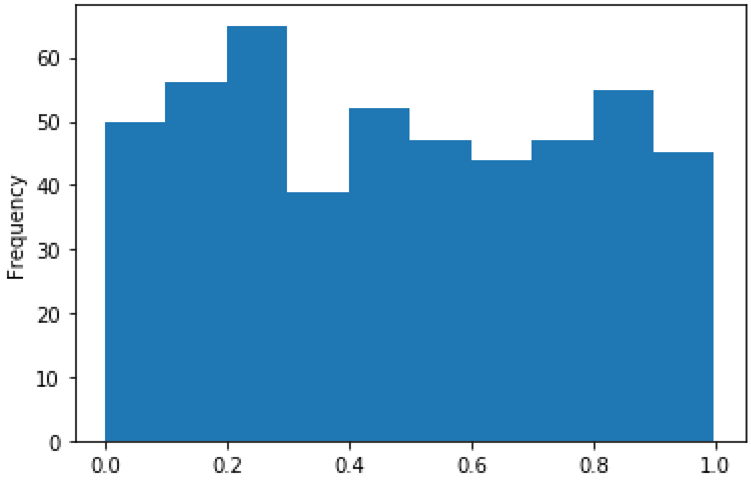
Out[10]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e32c718248>
```



** Create a histogram of the 'a' column.**

In [11]:
```
df3['a'].plot.hist()
```

Out[11]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e32d706548>
```

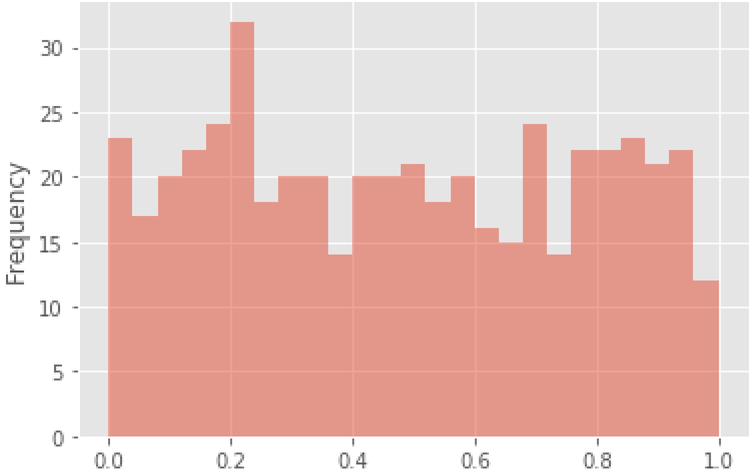


** These plots are okay, but they don't look very polished. Use style sheets to set the style to 'ggplot' and redo the histogram from above. Also figure out how to add more bins to it.***

```
In [12]: plt.style.use('ggplot')
```

```
In [13]: df3['a'].plot.hist(alpha=0.5,bins=25)
```

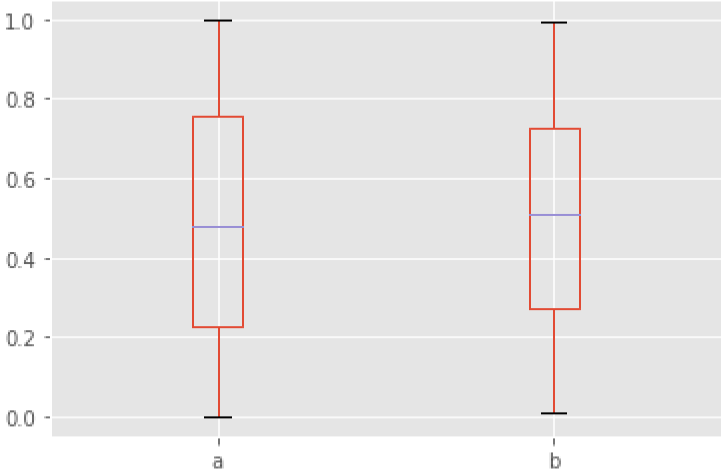
```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x1e32d7a6c08>
```



** Create a boxplot comparing the a and b columns.**

```
In [14]: df3[['a','b']].plot.box()
```

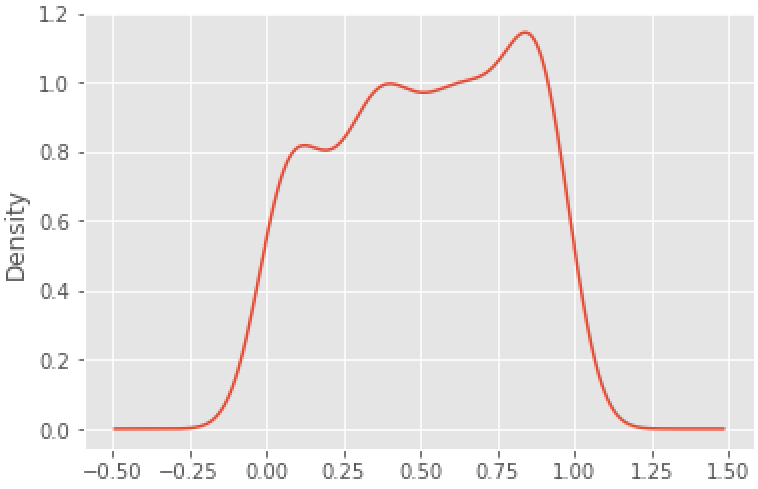
```
Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x1e32d81a988>
```



** Create a kde plot of the 'd' column **

```
In [15]: df3['d'].plot.kde()
```

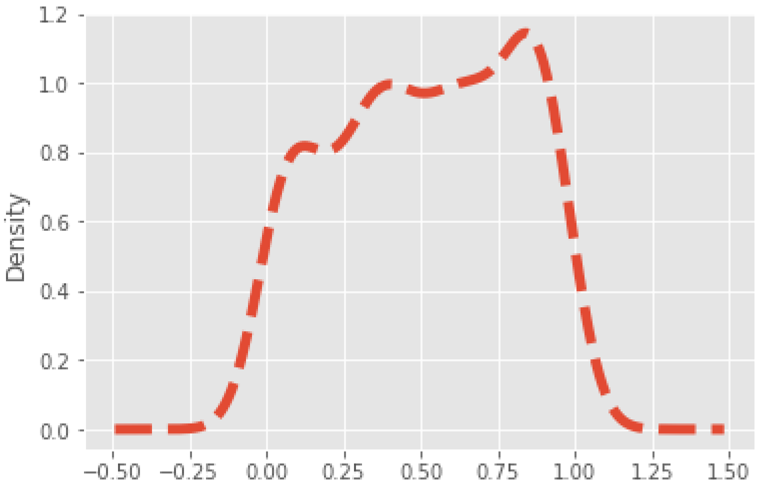
```
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x1e32d8e7148>
```



** Figure out how to increase the linewidth and make the linestyle dashed. (Note: You would usually not dash a kde plot line)**

```
In [16]: df3['d'].plot.density(lw=5,ls='--')
```

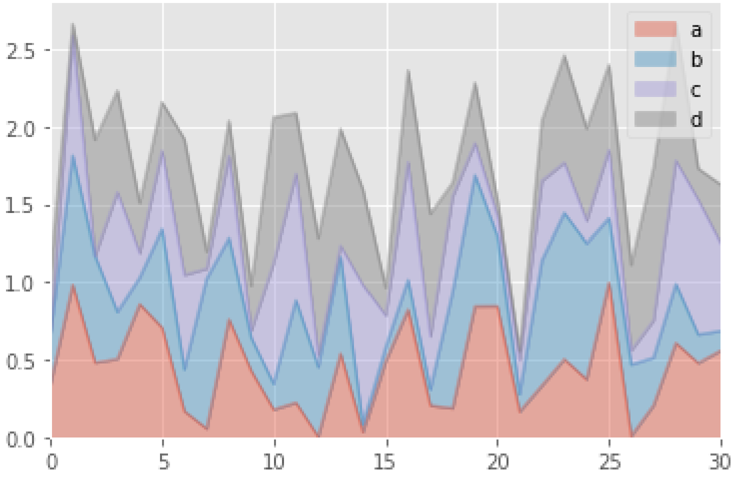
```
Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x1e32f3f2a88>
```



** Create an area plot of all the columns for just the rows up to 30. (hint: use .ix).**

In [22]:
```
df3.loc[0:30].plot.area(alpha=0.4)
```

Out[22]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1e32f7516c8>
```

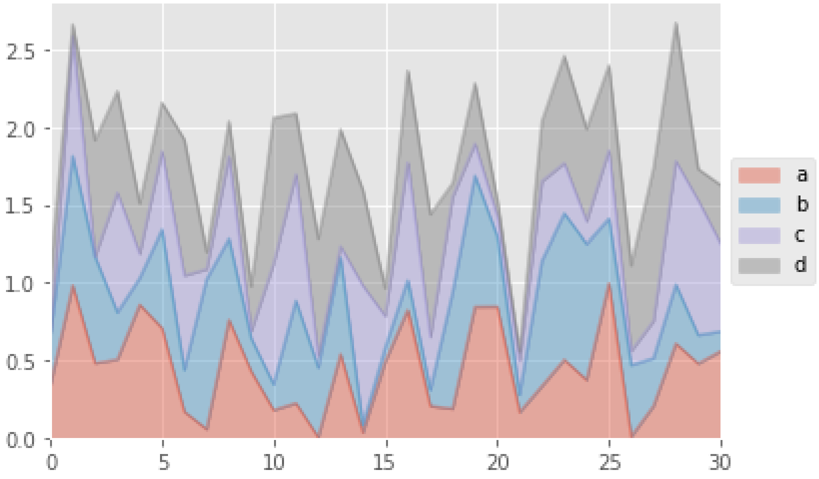


Bonus Challenge!

Note, you may find this really hard, reference the solutions if you can't figure it out! ** Notice how the legend in our previous figure overlapped some of actual diagram. Can you figure out how to display the legend outside of the plot as shown below?**

** Try searching Google for a good stackoverflow link on this topic. If you can't find it on your own - [use this one for a hint. (http://stackoverflow.com/questions/23556153/how-to-put-legend-outside-the-plot-with-pandas)](http://stackoverflow.com/questions/23556153/how-to-put-legend-outside-the-plot-with-pandas)**

In [21]:
```
f = plt.figure()
df3.loc[0:30].plot.area(alpha=0.4,ax=f.gca())
plt.legend(loc='center left', bbox_to_anchor=(1.0, 0.5))
plt.show()
```



Great Job!